Article

News Schemes for Activity Recognition Systems Using PCA-WSVM, ICA-WSVM, and LDA-WSVM

M'hamed Bilal Abidine * and Belkacem Fergani

Laboratoire d'Ingénierie des Systèmes Intelligents et Communicants, Faculty of Electronics and Computer Sciences, University of Science and Technology Houari Boumediene (USTHB), 32, El Alia, Bab Ezzouar, 16111 Algiers, Algeria; E-Mail: bfergani@gmail.com

* Author to whom correspondence should be addressed; E-Mail: mabidine@usthb.dz; Tel.: +213-551-96-42-64.

Academic Editor: Ahmed El Oualkadi

Abstract: Feature extraction and classification are two key steps for activity recognition in a smart home environment. In this work, we used three methods for feature extraction: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA). The new features selected by each method are then used as the inputs for a Weighted Support Vector Machines (WSVM) classifier. This classifier is used to handle the problem of imbalanced activity data from the sensor readings. The experiments were implemented on multiple real-world datasets with Conditional Random Fields (CRF), standard Support Vector Machines (SVM), Weighted SVM, and combined methods PCA+WSVM, ICA+WSVM, and LDA+WSVM showed that LDA+WSVM had a higher recognition rate than other methods for activity recognition.

Keywords: activity recognition; principal component analysis; independent component analysis; linear discriminant analysis; weighted support vector machines

1. Introduction

Activity recognition is one of the most important tasks in pervasive computing applications [1–4]. Research in human activity recognition is aimed to determine a human user's activity, such as cooking, brushing, dressing, sleeping, and so on. Therefore, different types of sensors have been used to sense user's activities in smart environments.

Sensor data collected needs to be analyzed using machine learning and pattern recognition techniques [5,6] to determine which activities is taking place by the dweller. As for any pattern recognition task, the keys to successful activity recognition are: (i) appropriately designed feature extraction of the sensor data; and (ii) the design of suitable classifiers to infer the activity. The learning of such models is usually done in a supervised manner and requires a large annotated dataset recorded in different settings [1–3].

The existing activity recognition algorithms suffer from two problems: the non informative of the feature space and the imbalanced data result in a degradation of the performance of activity recognition. Thus, feature extraction [7] preprocessing steps exist to extract a subset of new features from the original set by providing a better selection of relevant features of high-dimensional data, as well as high discrimination between classes. In this paper an attempt has been made to study three feature extraction methods, which are Principal Component Analysis (PCA) [8], Independent Component Analysis (ICA) [9], and Linear Discriminant Analysis (LDA) [10], and their relevance to improve the classification accuracy of the existing activity recognition systems.

Another problem affecting the performance of an algorithm's activity classification is the imbalanced data [11,12]. Activity recognition datasets are generally imbalanced, meaning certain activities occur more frequently than others (e.g., sleeping is generally done once a day, while toileting is done several times a day). This can negatively influence the learning process due to the known effect of minority class, which, in turn, imbalances the outcome, and may yield disastrous consequences for human activity recognition systems. This motivates extensive research that aims to improve the effectiveness of SVM on imbalanced classification in the activity recognition field [13,14]. Especially, approaches for addressing the imbalanced training-data problem can be categorized into two main streams: data processing approach and algorithmic approach [15–17].

The first approach is to preprocess the data either randomly or intelligently, by using undersampling the majority instances [16] or oversampling the minority instances [15]. In this paper, we consider the algorithmic approach in the following because it keeps all the information and does not change the distribution of training data. The solutions of this approach include the cost-sensitive learning [18,19] that treats different misclassifications using the weights assigned to data in order to pursue a high accuracy of classification.

Our paper addresses these issues and contributes on the following topics. Firstly, we have presented new schemes using PCA+WSVM, ICA+WSVM, and LDA+WSVM to recognize activities of daily living from binary sensor data. The Weighted Support Vector Machine (WSVM) [9] was employed to handle the imbalanced classification data problem using three methods, independently, for feature extraction: PCA, ICA, and LDA. Secondly, the proposed approaches are assessed and compared with the Conditional Random Fields (CRF) [20], the standard SVM, and Weighted SVM methods. Especially, CRF has recently gained popularity in the activity recognition field [1,3]. The experiments were implemented on multiple annotated real-world datasets from sensor readings in different houses [21,22].

2. Proposed Strategy Based Activity Recognition System

Despite its popularity in machine learning, the SVM technique has not been extensively used in activity recognition studies as pointed out in [23–26]. However, by the high accuracy rates obtained in other contexts, this would suggest possible success in activity recognition. Nevertheless, it is overwhelmed by the majority class instances in the case of imbalanced datasets. The weighted Support Vector Machine (WSVM) technique has been suggested as a candidate solution for such a purpose because it uses an efficient training approach that will improve its ability to learn from a large or imbalanced data set and, therefore, improve the performances of multi-class classifier SVM.

In this paper, a new activity recognition scheme is proposed; the WSVM method was applied for imbalanced classification using three methods, independently, for feature extraction: PCA, ICA, and LDA, as shown in Figure 1. PCA aims to eliminate the redundancy information. ICA estimates components as statistically independent as possible. LDA improves the separability of samples in the subspace and extracts LDA features. Then, these transformed (lower-dimensional space) datasets by each feature extraction method will be used for learning and testing a WSVM classifier. The outcome of the trained WSVM will then be used to process a new observations during the testing phase, where the associated activities of daily living class will be predicted.
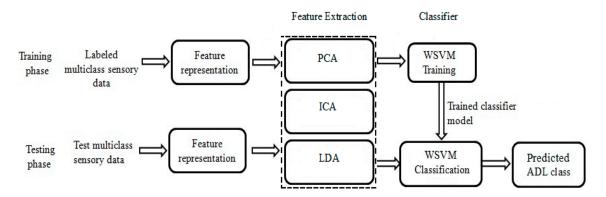

Figure 1. Scheme of the proposed strategy based activity recognition system.

2.1. Feature Extraction Methods

Suppose $X = \{x_i, i = 1, 2..., m\}$ are sets of training data with $x_i \in R^n$, *m* is the total of samples, *n* is sample's feature dimension, *N* is the total of classes. Projected sample is: $x_i \in R^p$ $(p < n)$.

2.1.1. Principal Component Analysis (PCA)

Principal component analysis [8] is a projection-based technique that approximates the original data with lower dimensional feature vectors through the construction of uncorrelated principal components that are a linear combination of the original variables. However, PCA is ignorant of the class labels attached to the data, so a good class separation in the direction of the high variance principal components is not guaranteed [8]. The main process of PCA is as follows.

In PCA, data matrix $X \in R^{m*n}$ are first centered $x \Leftarrow x - \bar{x}$ with $\bar{x}$ is the mean of the samples. Then PCA diagonalizes the covariance matrix as

$$Cov_{(X)} = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x})^T \text{ with } \bar{x} = \frac{1}{m}\sum_{i=1}^{m}x_i \tag{1}$$

This problem leads to solve the eigenvalue equation

$$\lambda V = Cov_{(X)} V, \|V\| = 1 \tag{2}$$

where $V = [v_1, v_2, \ldots, v_i]$, $(i = 1, \ldots, n)$ is the $n \times n$ matrix containing n eigenvectors and λ is an $n \times n$ diagonal matrix of eigenvalues of the covariance matrix. In Equation (2), each n dimensional eignvector v_i corresponds to the ith eigenvalue λ_i. The variance in any direction v_i can be measured by dividing the associated eigenvalue λ_i by the sum of the n eigenvalues. The first p principal components are selected as principal components which will be used for classification when their accumulative contributive rate is:

$$w = \frac{\sum_{j=1}^{p}\lambda_j}{\sum_{j=1}^{n}\lambda_j} \geq Threshold = 0.85 \tag{3}$$

2.1.2. Independent Component Analysis (ICA)

The most commonly used method for generating spatially-localized features is independent component analysis (ICA) to produce basis vectors that are statistically independent (not just linearly decorrelated, as with PCA) [9]. The algorithm works on the principle of minimizing mutual information between the variables; minimizing mutual information is the correct criteria for judging independence. Additionally, minimizing mutual information is same as maximizing entropy.

The ICA model can also be written as:

$$U = WX \tag{4}$$

Based on the knowledge of informatics, negentropy of U can be used as the criteria to estimate the independency of vectors, which is approximated by using the contrast function [27]:

$$J_G(w_i) = [E\{G(w_i^T X)\} - E\{G(V)\}]^2 \tag{5}$$

where V is the standardized Gaussian random variable (zero mean and unit variance). G is a non-quadratic function, the commonly used G can be:

$$G_1(u) = \frac{1}{\alpha_1}\log\cosh\alpha_1 u \tag{6}$$

$$G_2(u) = -\exp(-u^2/2) \tag{7}$$

where $1 \leq \alpha_1 \leq 2$ is some suitable constant.

Maximizing formula in Equation (5) leads to estimating w_i by

$$w_i^* = E\{Xg(w_i^T X)\} - E\{g'(w_i^T X)\}w_i \tag{8}$$

$$w_i = \frac{w_i^*}{\|w_i^*\|} \tag{9}$$

where w_i^* is a new estimated value of w_i. g and g' are respectively the first and second derivatives of G. Based on the maximal negentropy principal, the whole matrix W can be computed by maximizing the sum of one-unit contrast function and taking into account the constraint of decorrelation [27]. In practice, ICA can often uncover disjoint underlying trends in multi-dimensional data.

2.1.3. Linear Discriminant Analysis (LDA)

The aim of LDA is to find the optimal projection matrix $W_{opt} \in R^{n \times p}$ using the Fisher criterion below, to find the maximum of ratio of between-class scatter S_B to the within-class scatter S_W of the projected samples:

$$J(W_{opt}) = \text{argmax} \frac{\left\| W^T S_B W \right\|}{\left\| W^T S_W W \right\|} \tag{10}$$

where the between and within class covariance S_B and S_W are defined as:

$$S_B = \sum_{i=1}^{N} p_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \tag{11}$$

$$S_W = \frac{1}{m} \sum_{i=1}^{N} \sum_{x \in C_i} (x - \bar{x}_i)(x - \bar{x}_i)^T \tag{12}$$

where $p_i = m_i / m$ is *priori* probability of each class, m_i is the number of training samples of the ith class, $\bar{x}_i$ is the mean of the ith class, and $\bar{x}$ is the overall mean vector.

To maximize (10), the optimal W_{opt} are the eigenvectors associated with the largest eigenvalues of the following generalized eigenvalue problem:

$$S_B w_i = \lambda_i S_W w_i \tag{13}$$

The solution can be computed by solving the leading eigenvectors of $S_W^{-1} S_B$ that correspond to the eigenvalue λ_i. Then column vectors w_i are row vectors in the transformation matrix W. It should be noted that only those eigenvectors should be selected that correspond to eigenvalues carrying most of the energy, *i.e.*, the total dispersion. Another interesting property is that this transform decorrelates both S_B and S_W matrices. The rank of S_B is at most the N-1, and hence no more than this number of new features can be obtained.

2.2. Weighted Support Vector Machines (WSVM)

A SVM classifier is more insensitive to the problem of learning from imbalanced data. It considers a balanced training set using the same cost parameter C of different classes; this may generate suboptimal classification models. The SVM optimization primal problem is given as follows:

$$\min_{w,b,\xi} \quad 1/2 \, K(w,w) + C \sum_{i=1}^{m} \xi_i$$
$$subject \quad to \quad y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i, \, \xi_i \geq 0, i = 1,...,m \tag{14}$$

The Weighted Support Vector Machine (WSVM) was presented to deal with this problem by introducing two different cost parameters C^+ and C^- in the SVM optimization primal problem [5] for the majority classes ($y_i = +1$) and minority ($y_i = -1$), as given in Equation (15) below:

$$\min_{w,b,\xi} \quad 1/2.K(w,w) + C^+ \sum_{y_i=1}^{m+} \xi_i + C^- \sum_{y_i=-1}^{m-} \xi_i$$
$$subject \quad to \quad y_i(w^T \varphi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1,...,m \tag{15}$$

The dual optimization problem of WSVM with different constraints on α_i can be solved in the same way as solving the standard SVM optimization problem [5], which has the following dual form:

$$\max_{\alpha_i} \quad \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$\text{Subject to } \sum_{i=1}^m \alpha_i y_i = 0,$$
$$0 \leq \alpha_i \leq C_+, \text{ if } y_i = +1, \text{ and}$$
$$0 \leq \alpha_i \leq C_-, \text{ if } y_i = -1 \tag{16}$$

where m^+ and m^- are number of samples from +1 and −1 classes. C^+ and C^- are the cost parameters for positive and negative classes, respectively, to construct a classifier for multiple classes. They are used to control the trade-off between margin and training error. Some authors [19,28,29] have proposed adjusting different cost parameters for different classes of data, which effectively improves the low classification accuracy caused by imbalanced samples. Veropoulos *et al.* in [19] proposed to increase the trade-off associated with the minority class (*i.e.*, $C^- > C^+$) to eliminate the effect of class imbalance. However they did not suggest any guidelines to decide what the regularization factors should be. The coefficients are typically chosen as [30]:

$$C^+ = C \times w^+ \tag{17}$$

$$C^- = C \times w^- \tag{18}$$

When the two classes which request different sample size have the similar properties boundary (that is, the ratio of vectors supported by each class and their total sample size is equal, or these two classes have similar error rate), Hong Gunn Chew and others [30] took a detailed analysis for the reasons of classification accuracy caused by the size of the class in the SVM algorithm, and put forward the corresponding solutions. They obtained the following conclusions like this:

$$\frac{w^+}{w^-} = \frac{m^-}{m^+} \tag{19}$$

where C is the common cost coefficient for both classes in Equations (17) and (18), w^+ and w^- are the weights for +1 and −1 class respectively. In this paper, the weights are typically chosen as $w^+ = 1$ and $w^- = m^+/m^-$ for two-class WSVM. This criterion respects this reasoning that is to say that the tradeoff C^- associated with the smallest class is large in order to improve the low classification accuracy caused by imbalanced samples. The modified SVM algorithm would not tend to skew the separating hyperplane towards the minority class examples to reduce the total misclassifications as the minority class examples are now assigned with a higher misclassification cost.

For multiclass imbalanced data classification, we used different misclassification penalties per class. Typically the smallest class gets weighed higher. It allows the user to set individual weights for individual training examples, which are then used in WSVM training. We give the main ratio cost value C_i for each class i (1, …, N) in the function of the class prior probabilities $P(C_+)$ and $P(C_i)$ for the C_+ et C_i classes, respectively; it is given by:

$$C_i = C \times w_i \text{ where } w_i = \left[\frac{P(C_+)}{P(C_i)} \right]$$ (20)

We estimate each class prior probability $P(C_i)$ as the proportion of the number of samples in class i to the total number of training samples as follow:

$$p(C_i) = \frac{m_i}{\sum_i m_i}$$ (21)

Based on the above equation, the corresponding cost criterion in feature space can be given as follows:

$$C_i = C \times [m_+ / m_i], \; i = 1, ..., N$$ (22)

where m_+ is the number of samples of majority classes and m_i is the number of samples of the other classes. C is the common ratio misclassification cost factor of the WSVM. The search of the optimal value of the regularization parameter C is determined with the cross validation method. Where [.] is integer part of the quantity under square bracket. Notice that it always holds that $C_i \geq C$.

In this study, a software package LIBSVM [31] was used to implement the multiclass classifier algorithm. It uses the one-versus-one method (OVO) [5]. OVO method consists in constructing $N(N-1)/2$ classifiers and each one is trained on data from the two activity classes. When all $N(N-1)/2$ classifiers are constructed, a voting strategy is used for the test. The point is predicted in the class with the largest number of votes ("Max Wins" strategy). Chen *et al.* [32] discussed issues of using the same or different parameters for the $N(N-1)/2$ two-class problems. Their preliminary results show that both approaches give similar accuracy.

3. Experimental Setup and Results

3.1. Datasets

To evaluate the performance of our experimentations, we used different annotated datasets using different sensor networks in a pervasive environment [21,22]. The details of all the datasets are shown in Table 1. Each network was installed in a different home setting and was composed by a different number of sensors nodes. These sensors were installed in everyday objects such as doors, cupboards, refrigerator, and toilet flush to record activation/deactivation events (opening/closing events) as the subject carried out everyday activities. The sensor data were labeled using different annotation methods.

Table 1. House settings description.

Houses	TK26M	TK57M	TAP30F	TAP80F
Age	26	57	30	80
Gender	Male	Male	Female	Female
Annotation	Bluetooth headset	Handwritten diary	PDA	PDA
Duration	28 days	18 days	16 days	14 days
Sensors	14	21	77	84

A list of activities that were annotated for all datasets with the number of observations of each activity can be found in Table 2. Any period of time at which no activity took place was labelled "Idle". This table clearly shows how some activities occur very frequently (e.g., "toileting"), while others that occur less frequently have a longer duration (e.g., "leaving" and "sleeping"). Therefore, the datasets suffer from a severe class imbalance problem due to the nature of the data.

Table 2. Annotated list of activities for each house and the number of observations of each activity. The bold letters represent each activity.

TK26M	TK57M	TAP30F	TAP80F
Idle$_{(4627)}$	**Id**le$_{(2732)}$	**Id**le$_{(19025)}$	**Id**le$_{(17673)}$
Leaving$_{(22617)}$	**Lea**ving$_{(11993)}$	**Lea**ving$_{(87)}$	**To**ileting$_{(630)}$
Toileting$_{(380)}$	**Ea**ting$_{(376)}$	**To**ileting$_{(776)}$	Take **medication**$_{(185)}$
Showering$_{(265)}$	**To**ileting$_{(243)}$	**Ba**thing$_{(459)}$	Prep.**breakfast**$_{(466)}$
Sleeping$_{(11601)}$	**Show**ering$_{(191)}$	**Gro**oming $_{(484)}$	Prep.**lunch**$_{(843)}$
Breakfast$_{(109)}$	Brush **t**eeth$_{(102)}$	**Dre**ssing$_{(149)}$	Prep.**d**inner$_{(506)}$
Dinner$_{(348)}$	**Sha**ving$_{(67)}$	Prep.**breakfast**$_{(233)}$	Prep.**snack**$_{(320)}$
Drink$_{(59)}$	**Sleep**ing$_{(7738)}$	Prep.**lunch**$_{(676)}$	Washing **dishes**$_{(328)}$
	Dressing$_{(112)}$	Prep.**dinner$_{(178)}$**	Watching **TV**$_{(717)}$
	Medication$_{(16)}$	Prep.**snack**$_{(137)}$	Listen **music$_{(1100)}$**
	Breakfast$_{(73)}$	Preparing a beverage $_{(165)}$	
	Lunch$_{(62)}$	Washing **dishes**$_{(68)}$	
	Dinner$_{(291)}$	**Clea**ning$_{(186)}$	
	Snack$_{(24)}$	Doing **laundry$_{(246)}$**	
	Drink$_{(34)}$		
	Relax$_{(2435)}$		

3.2. Setup

The models were validated by splitting the original data into a test and a training set using a "Leave One Day Out cross validation" approach, retaining one full day of sensor readings for testing and using the remaining sub-samples as training data. The process is then repeated for each day and the average performance measure reported.

Sensor outputs are binary and represented in a feature space which is used by the model to recognize the activities performed. The vector contained one entry for each sensor, two-state sensors 0 or 1 are used and the features are the states of all sensors. The raw sensor representation uses the sensor data in the same way it was received from the sensor network. The value is 1 when the sensor is

active and 0 otherwise. We do not use the raw sensor data representation as observations; instead we use the combining "Change point" and "Last" representation which have been shown to give much better results in activity recognition [3].

In learning imbalanced data, the overall classification accuracy is not considered an appropriate measure of performance. Due to the fact that, in our case, we evaluate the models using F-Measure, a measure that considers the correct classification of each class is equally important. It is calculated from the precision and recall scores. We are dealing with a multi-class classification problem and therefore define the notions of true positive (TP), false negatives (FN), and false positives (FP) for each class separately. With highly-skewed data distribution, the overall accuracy metric at (23) is not sufficient anymore. It does not take into account differences in the frequency of activities. These measures are calculated as follows:

$$Accuracy \ = \frac{\sum_{i=1}^{N} TP_i}{Total} \times 100\% \tag{23}$$

$$Precision \ = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{TP_i}{TP_i + FP_i} \right] \times 100\% \tag{24}$$

$$Recall \ = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i} \times 100\% \tag{25}$$

$$F - Measure \ = \frac{2.Precision.Recall}{Precision + Recall} \times 100\% \tag{26}$$

3.3. Results

In our experiments, the SVM algorithm is tested with a LibSVM implementation [31]. It was used to implement the one-versus-one multiclass classifier [5]. We used the radial basis kernel function as follows: $K(x, y) = \exp\left(\frac{-1}{2\sigma^2} \|x - y\|^2 \right)$. Firstly, we optimized the SVM hyper-parameters (σ, C) for all training sets in the range (0.1–2) and [0.1, 1, 10, 100], respectively, to maximize the class accuracy of the leave-one-day-out cross validation technique. The best pair parameters (σ_{opt}, C_{opt}) = (1.7, 1), (2, 1), (1.4, 1), and (1.2, 1) are used for the datasets TK26M, TK57M, TAP30F, and TAP80F respectively. Then, locally, we optimized the cost parameter C_i, adapted for each activity class by using WSVM classifier with the common cost parameter is fixed $C = 1$, see Tables 3–6.

Table 3. Selection of the weights w_i using TK26M dataset.

Activity	Id	Le	To	Sh	Sl	Br	Di	Dr
w_i	5	1	61	88	2	216	73	419

Table 4. Selection of the weights w_i using TK57M dataset.

Activity	Id	Le	Ea	To	Sho	B.t	Sha	Sl	Dre
	4	1	32	50	63	118	179	2	107
w_i	**Me**	**Br**	**Lu**	**Di**	**Sn**	**Dri**	**Re**	-	-
	749	164	193	41	500	375	5	-	-

Table 5. Selection of the weights w_i using TAP30F dataset.

Activity	Id	Le	To	Ba	Gr	Dr	P.b	P.l	P.d
	1	220	24	40	38	126	82	28	101
w_i	**P.s**	**P.b**	**W.d**	**Cl**	**D.l**	-	-	-	-
	131	128	307	96	73	-	-	-	-

Table 6. Selection of the weights w_i using TAP80F dataset.

Activity	Id	To	T.m	P.b	P.l	P.d	P.s	W.d	W.TV	L.m
w_i	1	30	92	38	21	36	72	53	32	17

We reported in Figures 2 and 3 the selected features using PCA and LDA for all datasets. The summary of the performance measures obtained for all classifiers are presented in Table 7. For CRF results on these datasets, refer to [3,33,34]. ICA differs from PCA in the fact that the low-dimensional signals do not necessarily correspond to the directions of maximum variance. We start with the first independent component and keep increasing the number until the cross-validation error reduces.

After the selection of the best parameters, we evaluated the performance of different algorithms using appropriate metrics for imbalanced classification. The classification results for CRF, SVM, WSVM, PCA+WSVM, ICA+WSVM, and LDA+WSVM are summarized in Table 7 below.
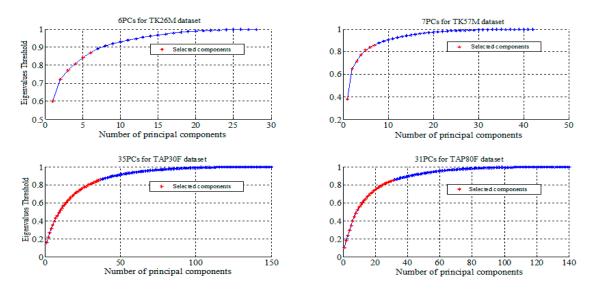

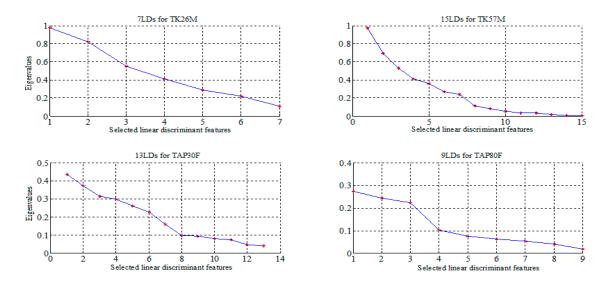**Figure 2.** Feature selection by Principal Component Analysis (PCA).

Figure 3. Feature selection by Linear Discriminant Analysis (LDA).

Table 7. Recall (Rec.), Precision (Prec.), F-measure (F), and Accuracy (Acc.) results for all methods. The values are percentages.

Dataset	Classifier	Rec.	Prec.	F	Acc.
TK26M	CRF [3]	70.8	74.4	72.5	95.6
	SVM	61.8	73.3	67.0	95.5
	WSVM	72.8	74.6	73.7	92.5
	PCA+WSVM	71.5	71.5	71.5	91.2
	ICA+WSVM	71.2	73.3	72.2	92.7
	LDA+WSVM	77.0	78.4	77.7	93.5
TK57M	CRF [33]	30.0	36.0	33.0	78.0
	SVM	35.6	34.9	35.2	80.8
	WSVM	40.8	37.8	39.2	77.1
	PCA+WSVM	36.5	34.2	35.3	76.9
	ICA+WSVM	36.2	38.1	37.1	76.6
	LDA+WSVM	42.3	39.8	41.0	77.2
TAP30F	CRF [34]	26.3	31.9	28.8	83.7
	SVM	22.3	34.0	26.9	83.3
	WSVM	30.8	30.6	30.7	23.8
	PCA+WSVM	32.1	31.6	31.8	20.8
	ICA+WSVM	30.4	28.7	29.5	21.7
	LDA+WSVM	38.2	52.9	44.3	33.8
TAP80F	CRF [34]	27.1	29.5	28.2	77.2
	SVM	15.2	30.0	20.1	75.6
	WSVM	29.2	29.4	29.3	28.7
	PCA+WSVM	29.6	29.4	29.5	22.4
	ICA+WSVM	26.5	27.9	27.2	22.1
	LDA+WSVM	38.7	45.7	41.9	28.7

This table shows that LDA+WSVM method gives a clearly better F-measure performance, while CRF and SVM methods perform better in terms of accuracy for all datasets. As can be noted in this

table, LDA outperforms PCA and ICA for recognizing activities with a WSVM classifier for all datasets. The PCA+WSVM method improves the classification results compared to CRF, SVM, WSVM, and ICA+WSVM for the TAP30F and TAP80F datasets, compared to other datasets.

The Figures 4 and 5 give the classification results in terms of the accuracy measure for each activity with WSVM, PCA+WSVM, ICA+WSVM, and LDA+WSVM methods.

In Figure 4, for WSVM, PCA+WSVM, LDA+WSVM models, the minority activities "Toileting", "Showering", and the kitchen activities "Breakfast" and "Drink" are significantly better detected, compared to other methods. LDA+WSVM is an effective method for recognizing activities. The majority activities are better for all methods, while the "Idle" activity is more accurate for the LDA+WSVM method.
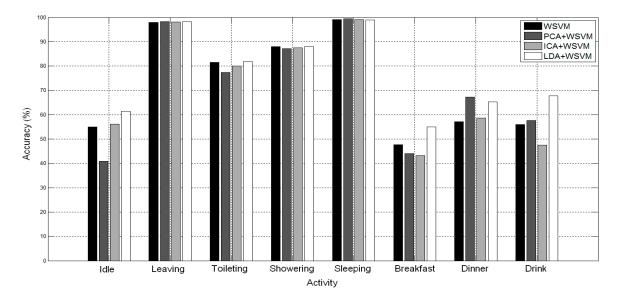

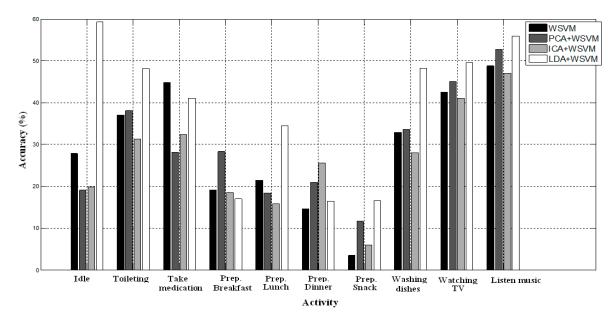**Figure 4.** Accuracy for each activity on TK26M dataset.

Figure 5. Accuracy for each activity on TAP80F dataset.

We can see in Figure 5 that the minority activities ("Toileting", "Washing dishes", "Watching TV", "Listen music", and the kitchen activities "Prep.Lunch", "Prep.Snack") are better recognized with LDA-WSVM. Additionally, the kitchen activities perform worst for all datasets. They are, in general, hard to recognize but they are better recognized with LDA-WSVM compared to others methods.

3.4. Discussion

Based on the experiments carried out in this work, a number of conclusions can be drawn. Using experiments on large real-world datasets, we showed the F-measure obtained with TK26M dataset is better compared to other datasets for all recognition methods because the TK57M, TAP30F, and TAP80F datasets include more activity classes. We supposed that the use of a hand-written diary in the TK57M house and PDA in TAP30F and TAP80F houses for annotating data is less accurate than using the Bluetooth headset as in TK26M house. For the TK26M dataset, a Bluetooth headset was used which communicated with the same server the sensor data was logged on. This means the timestamps of the annotation were synchronized with the timestamps of the sensors. In TK57M activity diaries were used, this is more error-prone because times might not always be written down correctly and the diaries have to be typed over afterwards.

In this section, we explain the difference in terms of performance between different recognition methods for imbalanced dataset. Our experimental results show that WSVM and LDA+WSVM methods work better for classifying activities; they consistently outperform the other methods in terms of the accuracy of the minority classes. In particular, LDA-WSVM is the best classification method for all datasets because the LDA method is more adapted for the features reduction in the datasets with consideration the discrimination between classes.

PCA-WSVM outperforms CRF, SVM, WSVM, and ICA-WSVM for TAP30F and TAP80F datasets. In other datasets ICA-WSVM surpasses PCA-WSVM. We conclude that the PCA method is more adapted for the features extraction in the datasets with large features vectors.

A multiclass SVM classifier does not take into consideration the differences (costs) between the class distributions during the learning process and optimizes with the cross-validation research the same cost parameter *C* for all classes. Not considering the weights in SVM formulation affects the classifiers' performances and favors the classification of majority activities ("Idle", "Leaving" and "Sleeping"). Although WSVM, including the individual setting of parameter *C* for each class, is significantly more effective than CRF and SVM methods, WSVM is not efficient compared to LDA+WSVM. The LDA method significantly improves the performance of the WSVM classifier. Thus, it follows that LDA-WSVM can be made more robust for classifying human activities.

The recognition of the minority activities in TK26M as "Toileting", "Showering", "Breakfast" "Dinner", and "Drink" is lower compared to "Leaving" and "Sleeping" activities. This is mainly due to the fact that the minority activities are less represented in the training dataset. However, the activities "Idle" and the three kitchen activities gave the worst results compared to the others activities. Most confusion occurs between the "Idle" activity and the kitchen activities. In particular, the "Idle" is one of the most frequent activities but is usually not a very important activity to recognize. It might, therefore, be preferable to lose accuracy on this activity if it allows a better recognition of minority classes.

The kitchen activities are food-related tasks, they are worst recognized for all methods because most of the instances of these activities were performed in the same location (kitchen) using the same set of sensors. In other words, it is observed that groups of similar activities are more separable if performed in different locations. For example, "Toileting" and "Showering" are more separable because they are in two different locations in the TK26M dataset. Therefore, the location of the sensors is of great importance for the performance of the recognition system.

4. Conclusions

In this paper, we have proposed a combination of PCA, ICA, and LDA methods and a Weighted SVM prediction model to recognize activities of daily living from home environments using a network of binary sensors. The proposed scheme shows two merits:

(1) After the feature extraction step with PCA, ICA, and LDA, the most significant components to the set of features extracted are obtained, the training set is reduced, and the prediction accuracy is improved.
(2) The multi-class Weighted SVM classifier, as the latter processor, has good generalization performance in imbalanced human activity datasets.

Experimental results show that the LDA-WSVM learning method produces interesting results for activity recognition success. This model is effective to classify multiclass sensory data over techniques such as CRF, SVM, WSVM, PCA-WSVM, and ICA+WSVM. In all datasets LDA-WSVM has the highest F-measure metric, while CRF and SVM models produced high accuracy. This is due to the fact that CRF and SVM are more sensitive to overfitting on a dominant class than other methods. Finally, we observed that differences in the layout of houses and the way a dataset was annotated used for training the models can greatly affect the performance in activity recognition. In this work, we have used the offline inference. The activities could only be inferred when a full day has passed. It would also be interesting to perform the LDA+WSVM method in online inference, which is significantly harder; however, it is necessary for specific applications.

Author Contributions

M'hamed Bilal Abidine: Main writing and also analyzing and improving the proposed approach; Belkacem Fergani: Total supervision of the work, review, and comments. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Tapia, E.M.; Intille, S.S.; Larson, K. Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive Computing*, Proceedings of the Second International Conference,

PERVASIVE 2004, Linz/Vienna, Austria, 21–23 April 2004; Springer: Berlin/Heidelberg, Germany; Lecture Notes in Computer Science, Volume 3001; pp. 158–175.

2. Logan, B.; Healey, J.; Philipose, M.; Tapia, E.M.; Intille, S.S. A long-term evaluation of sensing modalities for activity recognition. In Proceedings of the 9th International Conference on Ubiquitous Computing (Ubicomp '07), Innsbruck, Austria, 16–19 September 2007; pp. 483–500.

3. Van Kasteren, T.L.M.; Noulas, A.; Englebienne, G.; Kröse, B.J. Accurate activity recognition in a home setting. In Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08), Seoul, Korea, 21–24 September 2008; pp. 1–9.

4. Fleury, A.; Vacher, M.; Noury, N. SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 274–283.

5. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 2000.

6. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.

7. Guyon, I.; Gunn, S.; Nikravesh, M.; Zadeh, L. *Feature Extraction: Foundations and Applications*; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2006.

8. Jolliffe, I.T. *Principal Component Analysis*; Springer: New York, NY, USA, 2002.

9. Comon, P. Independent component analysis—A new concept? *Signal Process.* **1994**, *36*, 287–314.

10. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning*; Springer: Heidelberg, Germany; New York, NY, USA, 2001.

11. Chawla, N.V. Data mining for imbalanced datasets: An overview. In *Data Mining and Knowledge Discovery Handbook*; Springer: New York, NY, USA, 2010; pp. 875–886.

12. Van Hulse, J.; Khoshgoftaar, T.M.; Napolitano, A. Experimental perspectives on learning from imbalanced data. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 935–942.

13. Cao, H.; Nguyen, M.; Phua, C.; Krishnaswamy, S.; Li, X. An Integrated Framework for Human Activity Classification. In Proceedings of the 14th ACM International Conference on Ubiquitous Computing (UbiComp '12), Pittsburgh, PA, USA, 5–8 September 2012; pp. 331–340.

14. Abidine, M.B.; Fergani, B.; Clavier, L. Importance-Weighted the Imbalanced data for *C*-SVM Classifier to Human Activity Recognition. In Proceedings of the 8th International Workshop on Systems, Signal Processing and their Applications (WOSSPA'13), Algiers, Algeria, 12–15 May 2013; pp. 330–335.

15. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Rtificial Intell. Res.* **2002**, *16*, 321–357.

16. Vilarino, F.; Spyridonos, P.; Vitrià, J.; Radeva, P. Experiments with SVM and stratified sampling with an imbalanced problem: Detection of intestinal contractions. In Proceedings of the 3rd ICAPR, Bath, UK, 22–25 August 2005; Springer: Berlin/Heidelberg, Germany; Lecture Notes in Computer Science, Volume 3687; pp. 783–791.

17. Akbani, R.; Kwek, S.; Japkowicz, N. Applying support vector machines to imbalanced datasets. In *Machine Learning: ECML'04*, Proceedings of the 15th European Conference on Machine Learning, Pisa, Italy, 20–24 September 2004; pp. 39–50.

18. Huang, Y.M.; Du, S.X. Weighted support vector machine for classification with uneven training class sizes. In Proceedings of the 4th IEEE International Conference on Machine Learning and Cybernetics, Guangzhou, China, 18–21 August 2005; pp. 4365–4369.

19. Osuna, E.; Freund, R.; Girosi, F. *Support Vector Machines: Training and Applications*; Technical Report A.I. Memo No. 1602; Massachusetts Institute of Technology: Cambridge, MA, USA, 1997.

20. Vail, D.L.; Veloso, M.M.; Lafferty, J.D. Conditional Random Fields for Activity Recognition. In Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007), Honolulu, HI, USA, 14–18 May 2007; doi:10.1145/1329125.1329409.

21. Van Kasteren, T.L.M. Datasets for Activity Recognition. Available online: http://sites.google.com /site/tim0306/ (accessed on 9 February 2012).

22. Tapia, E.M. Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors. Available online: http://courses.media.mit.edu/2004fall/mas622j/04.projects/home/ (accessed on 3 April 2013).

23. Abidine, M.B.; Fergani, B. Evaluating C-SVM, CRF and LDA classification for daily activity recognition. In Proceedings of the 2012 International Conference on Multimedia Computing and Systems (ICMCS), Tangier, Morocco, 10–12 May 2012; pp. 272–277.

24. Banos, O.; Damas, M.; Pomares, H.; Prieto, A.; Rojas I. Daily living activity recognition based on statistical feature quality group selection. *Expert Syst. Appl.* **2012**, *39*, 8013–8021.

25. Manosha Chathuramali, K.G.; Rodrigo, R. Faster Human Activity Recognition with SVM. In Proceedings of the 2012 IEEE International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 12–15 December 2012; pp. 197–203.

26. Palaniappan, A.; Bhargavi, R.; Vaidehi, V. Abnormal Human Activity Recognition Using SVM Based Approach. In Proceedings of the 2012 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, India, 19–21 April 2012; pp. 97–102.

27. Hyvärinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **2000**, *13*, 411–430.

28. Veropoulos, K.; Campbell, C; Cristianini, N. Controlling the Sensitivity of Support Vector Machines. In Proceedings of the International Joint Conference on AI, Stockholm, Sweden, 31 July–6 August 1999; pp. 55–60.

29. Cao, P.; Zhao, D.; Zaiane, O.R. An Optimized Cost-Sensitive SVM for Imbalanced Data Learning. In *Advances in Knowledge Discovery and Data Mining*; Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G., Eds.; Springer: New York, NY, USA, 2013; pp. 280–292.

30. Chew, H.G.; Crisp, D.J.; Bogner, R.; Lim, C.C. Target detection in radar imagery using support vector machines with training size biasing. In Proceedings of the 6th International Conference on Control, Automation, Robotics, and Vision (ICARCV), Singapore, Singapore, 5–8 December 2000; pp. 80–85.

31. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. Available online: http://www.csie.ntu.edu.tw/~cjlin-/libsvm/ (accessed on 22 April 2010).

32. Chen, P.H.; Lin, C.J.; Scholköpf, B. A tutorial on *nu*-Support Vector Machines. *Appl. Stoch. Models Bus. Ind.* **2005**, *21*, 111–136.

33. Van Kasteren, T.L.M.; Alemdar, H.; Ersoy, C. Effective Performance Metrics for Evaluating Activity Recognition Methods. In Proceedings of the ARCS 2011—24th International Conference on Architecture of Computing Systems, Comot, Italy, 24–25 February 2011; p. 10.

34. Van Kasteren, T.L.M.; Englebienne, G.; Kröse, B.J. An activity monitoring system for elderly care using generative and discriminative models. *Pers. Ubiquitous Comput*. **2010**, *14*, 489–498.